

# Multiple Coordinated Views for Searching and Navigating Web Content Repositories

Alexander Hubmann-Haidvogel\*

alexander.hubmann@modul.ac.at

Arno Scharl\*

arno.scharl@modul.ac.at

Albert Weichselbraun†

albert.weichselbraun@wu.ac.at

January 2009

## Abstract

The advantages and positive effects of multiple coordinated views on search performance have been documented in several studies. This paper describes the implementation of multiple coordinated views within the *Media Watch on Climate Change*, a domain-specific news aggregation portal available at [www.ecoresearch.net/climate](http://www.ecoresearch.net/climate) that combines a portfolio of semantic services with a visual information exploration and retrieval interface. The system builds contextualized information spaces by enriching the content repository with geospatial, semantic and temporal annotations, and by applying semi-automated ontology learning to create a controlled vocabulary for structuring the stored information. Portlets visualize the different dimensions of the contextualized information spaces, providing the user with multiple views on the latest news media coverage. Context information facilitates access to complex datasets and helps users navigate large repositories of Web documents. Currently, the system synchronizes information landscapes, domain ontologies, geographic maps, tag clouds and just-in-time information retrieval agents that suggest similar topics and nearby locations.

## 1 Introduction

This paper investigates automated methods to build large Web content repositories and visual means for searching and exploring these repositories. Semantic annotations are a fundamental part of this process, which is reflected in the major research questions underlying this paper: What are the most effective and scalable methods to identify significant phrases in general and geospatial references in particular? How can the system assess the phrases' relative importance in a given body of text? Is it possible to generate controlled vocabularies based on semantic associations identified within large document collections? How can we integrate and visualize semantic, geospatial and temporal annotations to facilitate Web search?

\*MODUL University Vienna, Department of New Media Technology, [www.modul.ac.at/nmt](http://www.modul.ac.at/nmt)

†Vienna University of Economics and Business, Research Institute for Computational Methods, [www.wu.ac.at/firm](http://www.wu.ac.at/firm)

This is a preprint of an article published in A. Hubmann-Haidvogel, A. Scharl, A. Weichselbraun. 2009. *Multiple Coordinated Views for Searching and Navigating Web Content Repositories. Information Sciences* 179 (12): 1813-1821.

The *Media Watch on Climate Change* ([www.ecoresearch.net/climate](http://www.ecoresearch.net/climate)) addresses these questions by integrating semantic Web and natural language processing technologies with a tightly coupled visual interface to convey context information. Figure 1 shows a screenshot of the current prototype, which aims to increase awareness and availability of environmental information by providing a comprehensive and continuously updated account of media coverage on climate change and related issues. The system aggregates, filters and visualizes environmental Web content from 150 Anglo-American news media sites.

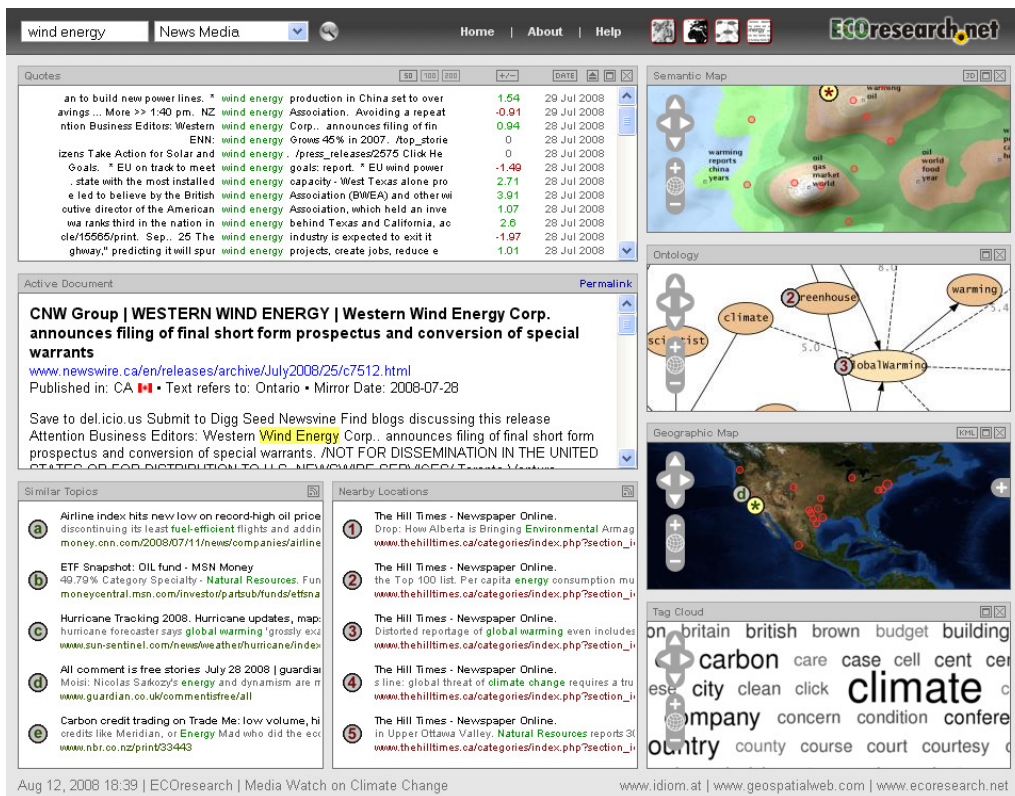


Figure 1: Screenshot of the *Media Watch on Climate Change* ([www.ecoresearch.net/climate](http://www.ecoresearch.net/climate))

The example query for 'wind energy' produces a list of relevant sentences including their sentiment and date of publication. On the right side of the screen, users can choose from several visualizations including information landscapes, geographic maps, domain ontologies and tag clouds (See Section 4.1: User Interface Design).

The prototype's visual user interface paves the way for more efficient ways of searching and browsing, while Web services and machine-readable semantic interfaces make the contextualized information space available for third-party applications - e.g., plug-ins for Web browsers that process microformats.

The remainder of this paper is organized as follows: Section 2 presents a literature review and highlights related research projects. Section 3 summarizes the methods employed and is structured according to the context dimension that each method is targeting (geospatial, semantic and temporal). Section 4 focuses on human consumption of the content repository by describing the prototype's interface design and specific implementation details. Section 5 focuses on machine consumption by outlining methods to provide data for third-party applications via a standardized Web Service interface. The paper concludes with a summary and outlook on possible future research avenues in Section 6.

## 2 Related Research

Improving information retrieval performance metrics like precision and recall does not necessarily translate into increased user satisfaction and better user performance for specific search tasks [2, 26]. Multiple coordinated views address this problem by structuring the search domain along several dimensions [22]. Users can navigate along these dimensions and access continuously updated context information.

The user interface of the *Media Watch on Climate Change* represents each dimension by at least one view (= portlet) in the user interface. User actions in any of the provided portlets trigger an immediate update of the context information provided by all other portlets. This is the essence of multiple coordinated views, which are also referred to as linked or tightly coupled views in the literature. The positive effects of multiple coordinated views on user performance have been documented in several studies [7, 17, 23]. They reduce search time, allow the detection of patterns, and aid users in choosing the next node to visit [6, 16]. Users of multiple coordinated views have been found to be more efficient in retrieving information and incorporating more information into their daily workflows compared to traditional search engines [19].

Coordinated multiple views provide effective navigation mechanisms for different data types. Baldonado and Kuchinsky [4] list guidelines for using multiple views and emphasize their potential to reveal correlations and disparities. Data displayed in multiple views also needs to be consistent; selecting objects in one view, for example, should automatically highlight the same objects in related views. Furthermore, the relations between different views need to be made apparent to the user, which can be accomplished by quickly propagating changes between the views.

Marchionini [15] describes the concept of exploratory search, combining querying and browsing datasets and using work in human-computer interaction and information retrieval to support users in their search tasks. *Snap Together* [18] is an early example in which users dynamically combine and bind different visualizations on-the-fly to customize their interface. *Oculus GeoTime* [12] integrates geospatial information and the temporal developments into a single 3D view. Brodbeck and Girardin [5] employ multiple coordinated views for analyzing geo-referenced, high-dimensional data sets. The *Media Watch on Climate Change* also falls into the group of geospatial-topical systems by providing both geospatial and semantic views for browsing large document sets within a single, coordinated interface.

## 3 Method

The *Media Watch on Climate Change* integrates technologies from natural language processing, information retrieval and ontology learning. It provides information retrieval agents and interactive visualizations that facilitate queries along multiple dimensions of context. The portal uses a contextualized information space as the primary knowledge base. To build this information space, the project drew upon the directories Newslink.org, Kidon.com, and ABYZNewsLinks.com to compile a list of the most influential news media sites - 42 from the United States and 72 from four other English-speaking countries: Canada, United Kingdom, Australia, and New Zealand. A Web crawler regularly mirrors these Web sites by following their hierarchical structure until reaching 50 megabytes of textual data. This limit helps compare sites of heterogeneous size and reduces the dilution of top-level information by articles found in lower hierarchical levels. A content filter then selects about 10,000 documents relevant to climate change and related issues from a total of about 300,000 documents that are being mirrored in weekly intervals.

If authors do not provide contextual attributes, *annotation services* attempt to tag and classify new objects. If contextual attributes are provided, the annotation services validate the manual entries, add missing information, and suggest changes if conflicting information is found. The annotation services of the current prototype provide metadata along three dimensions: (i) *geospatial* - distinguishing between source and target geography; (ii) *semantic* - computing significant phrases, keywords and assigning the most relevant concepts from a controlled vocabulary; and (iii) *temporal* - adding timestamps for the reported event, the

initial publication and subsequent revisions. Each element of the contextualized information space can be organized, indexed, searched and navigated along these dimensions.

The need for controlled vocabularies and shared meaning suggests that ontologies understood as shared conceptualizations within a specific domain [10] are going to play a key role in managing context information. Geo-ontologies encode geographical terms and relations such as containment, overlap and adjacency. Domain ontologies are used to organize topics and navigate through knowledge repositories. Spatially aware search engines use ontological knowledge for query term expansion and disambiguation, relevance ranking and Web resource annotation [1].

The presented approach is not supposed to generate a universally accepted domain ontology, but it should accurately represent knowledge contained within a specific corpus. Changes in sample composition or trends in news media coverage inevitably affect the generated ontology. This represents a promising research avenue for conducting comparative studies and investigating ontology evolution.

### 3.1 Geospatial Annotation

When analyzing geospatial context, it is important to distinguish between the *source geography*, which specifies the location of the publisher, and the *target geography*, which indicates the specific locations discussed in the Web document. The *Media Watch on Climate Change* uses the official headquarters of news media outlets as the source geography. Determining the target geography requires parsing and analyzing the Web document for geospatial references - a process that is usually referred to as geotagging. A promotional article describing the idyllic panoramic view from the “Marterle” (Austrian pilgrimage church at 1861 meters above sea level), for example, would be tagged with a geographic focus of Carinthia, Austria.

The geotagging component identifies geo-locations based on a geographic database (gazetteer) containing all relevant terms referring to a particular location. A focusing algorithm selects the most suitable locations for documents containing multiple possible geographic focuses. Based on the information found in the geo-database, a tree is built that aligns locations hierarchically (for example, Austria is a child of Europe, Vienna a child of Austria, etc.). The number of occurrences is only one indicator for the significance of the location. In the case of ambiguities (two or more locations referenced the same number of times), the location tree is used to derive the most suitable candidate. Factors like the location’s population also determine the outcome of the ranking process [3]. Depending on the specific application, the tagger can either return just the most relevant result including its probability, or a list of all possible matches.

Free public gazetteers like *GeoNames.org* provide detailed information about geographic entities, including their names in different languages, their geographic position (latitude and longitude), administrative divisions, populations and the location’s timezone in an easily parseable format. Refining these data yields the language- and region-specific database used in the geotagger. The gazetteer’s scope, coverage, granularity, balance and richness of annotation influence the outcome of the annotation process [13]. To keep the geotagger independent of the application and the language under consideration, we have implemented a generic mechanism for selecting a subset of gazetteer entries covering different regions and languages. A combination of ISO 639-1 language and ISO 3166 country codes is used to select the gazetteer suitable for the geotagging task (e.g., “de\_AT” and “en\_AT” for Austrian locations in German and English documents, respectively).

### 3.2 Semantic Annotation

The following section outlines the process of semantically annotating the content repository. Statistical methods identify significant phrases and keywords in the corpus. Automated ontology learning helps to structure semantic annotations and to create a domain-specific graph for navigating the Web content repository.

## Identifying Significant Phrases

The most significant phrases for a corpus are calculated using the log-likelihood ratio as described by Dunning [8]. This algorithm ranks bi- and trigrams according to the ratio between two probabilities: obtaining the observed word counts under the assumption that the terms in the n-gram are dependent on each other and under the assumption of being independent of each other.

For bigrams, the number of observed occurrences  $n_{ij}$  and the number of expected occurrences  $m_{ij}$  under the hypothesis of independent terms are calculated, and the log likelihood ratio is determined according to the following equation:

$$G^2 = 2 \sum_{i,j} n_{ij} \log\left(\frac{n_{ij}}{m_{ij}}\right). \quad (1)$$

When this method is generalized and applied to trigrams, four different models of independence between the three terms have to be taken into account. The final likelihood value is the lowest of the four calculated values. The resulting list of ranked phrases is stored together with statistical information such as term frequencies, which are used to calculate keywords as described in the following section.

## Calculating Keywords

The keywords for the current week are calculated using a chi-square test with Yates' correction for continuity [27], which compares the text published in the current week (= target corpus) to a reference corpus consisting of data from multiple previous weeks. The method eliminates common words such as prepositions, articles and pronouns, since they occur with equal relative frequency in both the target and the reference corpus. The expected frequency  $E_i$  is compared to the observed frequency  $O_i$ . The chi-square test with Yates' correction yields the significance of the deviation between the expected and observed frequencies.

$$\chi_{Yates}^2 = \sum_{i=1}^N \frac{(|O_i - E_i| - 0.5)^2}{E_i} \quad (2)$$

The calculation is typically applied to unigrams. Significant phrase detection (see previous section) yields frequency and likelihood values for multi-word terms. This allows extending the calculation to bi- and trigrams, which provides more descriptive keywords such as "Kyoto Protocol" and "Earth Day Network".

## Generating Controlled Vocabularies

Due to the sheer volume of data, using ontologies to structure large information spaces is difficult without an ontology extension system that suggests concepts and relations automatically. Figure 2 presents a conceptual view of the ontology extension architecture underlying the *Media Watch on Climate Change*.

A small set of terms obtained from domain experts or third-party sources is selected as seed ontology. An analyzer distributes the seed terms to evidence source plug-ins that suggest promising new concepts based on the document repository. Currently, the system acquires data from three evidence sources:

1. *Co-occurrence analysis* [20] at both the sentence level and the document level uses a threshold based on the co-occurrence significance to select 20 terms on each level.
2. *Trigger Phrases* [11] match a fragment of text that indicates a particular relation between terms (e.g. parent-child relation).
3. The *WordNet* lexical dictionary [9] is queried after disambiguating the seed ontology concepts using a vector space model.

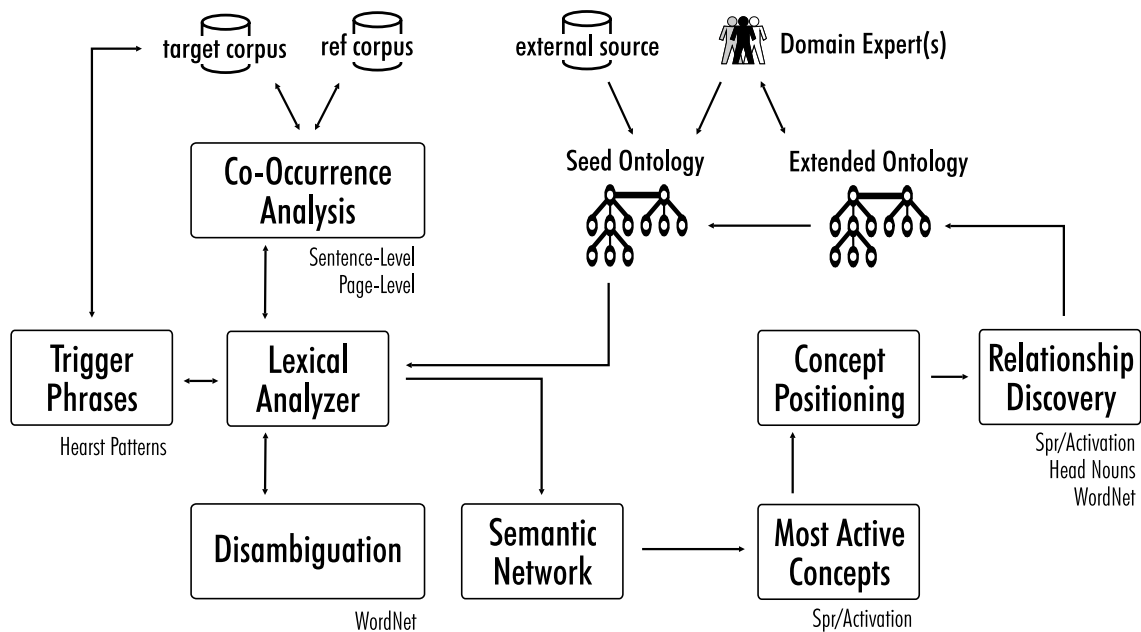


Figure 2: Ontology extension system architecture

The generated terms are then connected to a seed ontology via directed weighted links. Once the network is established, a spreading activation algorithm identifies the most relevant terms and suggests their incorporation into the seed ontology as described in Liu et. al [14]. Then the following heuristic is used to determine their relation to the concepts in the seed ontology: (i) WordNet, head noun analysis and additional rounds of spreading activation determine the new concepts' position within the ontology; (ii) Subsumption analysis together with WordNet and head noun analysis identify the relation type. To extract additional concepts for the ontology, another iteration of spreading activation is triggered. The center diagram of Figure 4 exemplifies such an automatically generated ontology. The lighter shading of the seed ontology concepts differentiates them from concepts added in several iterations of automated ontology learning.

### 3.3 Temporal Annotation

Snapshot analyses of Web content repositories fail to capture temporal effects, which require a more advanced and fine-grained approach. Ideally, Web crawlers only retrieve documents after registering a change in their content. The *Media Watch on Climate Change* mirrors well-defined samples of Web sites and Really Simple Syndication (RSS) feeds. The spider mirrors Web pages incrementally in daily, weekly and monthly intervals, using the "last-modified" header as specified in RFC 2616 to only retrieve documents whose date has changed since the previous visit. Continuously monitoring the RSS feeds defined in the sample list yields nearly instant representations of highly volatile content. To capture the temporal dimension of the contextualized information space, the spider stores the document's last modification date, the first time it was retrieved, and whether and when the document was removed from the server.

## 4 Prototype

One of the main research goals behind developing the *Media Watch on Climate Change* was the creation of a slim and high-performance framework for visualizing different views on contextualized content repositories. The emphasis on scalability lead to choosing the map paradigm for all visualizations. A Web Map Server (WMS) serves the image data and provides a common infrastructure for two-dimensional geographic maps, semantic maps, domain ontologies, and tag clouds. The UMN MapServer ([mapserver.gis.umn.edu](http://mapserver.gis.umn.edu)) provides the server-side functions. TileCache ([www.tilecache.org](http://www.tilecache.org)) adds a caching framework to increase query performance and reduce server load. On the client, OpenLayers ([www.openlayers.org](http://www.openlayers.org)) was chosen as a highly extensible library for displaying the maps.

### 4.1 User Interface Design

In the introductory section, Figure 1 shows a screenshot of the current prototype. On the left, the *Active Document* is displayed, including metadata such as its mirror date and source/target geography. Below, just-in-time information retrieval agents list documents referring to similar topics and nearby locations. Clicking on the related references extends the quoted text and clicking on the circular marker on the left activates that particular document.

The semantic and geographic maps facilitate access to the underlying knowledge base. The information landscape in the upper right window shows the semantic associations between documents. Peaks indicate clusters of documents about a popular topic, whereas valleys represent sparsely populated parts of the information space. Additional views such as a visual representation of an automatically generated domain ontology (see previous section) can be added to the display using toggle buttons above the map section. A geographic map shows the locations referred to in the listed documents. The tag cloud summarizes the repository's most prevalent topics, indicating the relative importance of a term by its font size. Clicking on the maximize buttons increases the size of the maps. The '3D' button of the semantic map opens a three-dimensional *Knowledge Planet* using the NASA World Wind Java SDK ([worldwind.arc.nasa.gov/java](http://worldwind.arc.nasa.gov/java)). The 'KML' button of the geographic map points towards a KML (Keyhole Markup Language) file that instructs geo-browsers such as NASA World Wind ([worldwind.arc.nasa.gov](http://worldwind.arc.nasa.gov)) and Google Earth ([earth.google.com](http://earth.google.com)) to display the 100 most relevant news items on a three-dimensional virtual globe.

Once the user enters a search query, an additional window displays quotes including the target term as 'concordances' (centering the target term and showing its immediate context in the various documents). Besides entering query terms, users can click on *any* position on the maps (not only on the document markers) to retrieve articles related to that particular location, topic or domain concept. The different views are therefore said to be 'tightly coupled': User actions in one window trigger an immediate update of all other displays. The continuous and synchronized display of several views on the contextualized information space allows rapid and reversible interaction.

### 4.2 View Synchronization

Figure 3 shows the system architecture for synchronizing multiple coordinated views. Using a simple observer pattern, each view registers itself to be notified of updates to the main view. Interaction in any coordinated view notifies the main view, which in turn notifies and triggers updates in all related views. Any view can trigger an update as a reaction to a user request, such as the selection of a related document in the similar topics list (dashed arrow). The main view is updated to show the selected document and retrieves new data from the Web server (dotted arrows). Updating the main document triggers updates for all related views (solid arrows), which fetch and display new data from the Web server. Data fetching is done using asynchronous background HTTP requests, avoiding full page reloads and giving the user instant updates and feedback.

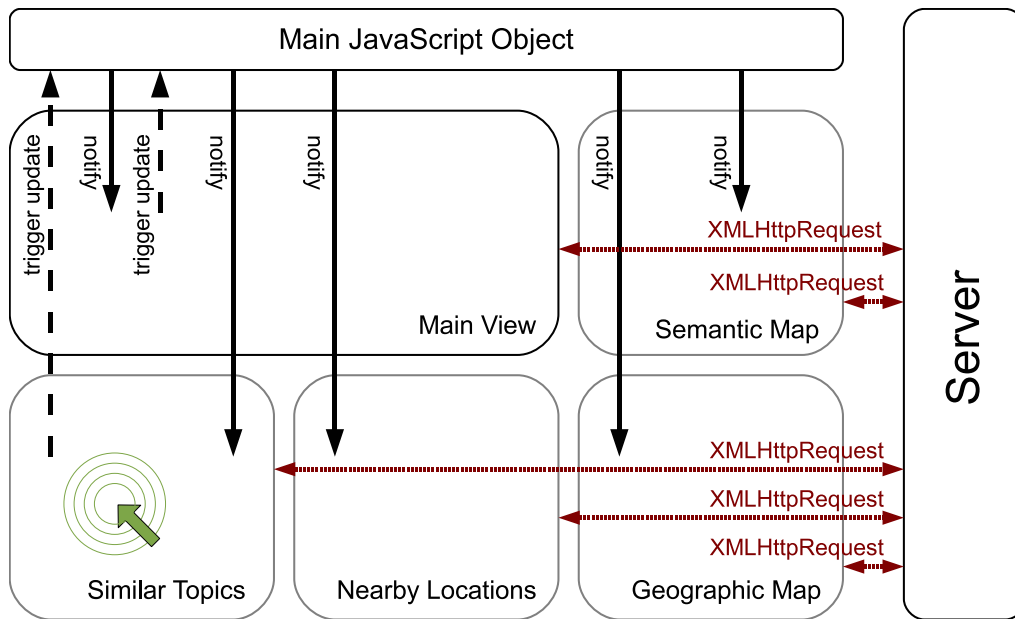


Figure 3: System architecture for synchronizing multiple coordinated views

### 4.3 Geospatial View

Users can select documents by clicking on a specific location on the geographic map and choose between different base layers including *NASA Blue Marble Next Generation* and *Google Maps Terrain*. A separate layer with document markers shows the target geography for the five documents most similar to the currently selected document. Additionally, the aggregated results of full text searches are displayed using colored circles, giving the user instant feedback about the location of all matching documents. The radius of a circle corresponds to the number of references to a particular location.

### 4.4 Semantic Views

Figure 4 compares three navigational aids based on semantic annotations: (i) a semantic map that positions documents according to their similarity on an artificial topography; (ii) a domain ontology that summarizes the domain’s most important concepts and their relations; and (iii) a tag cloud that contains keywords extracted from the contextualized information space.

*Semantic Map.* The artificial landscape is created by pre-clustering all documents in the corpus and using a force-directed placement algorithm where similar documents attract each other over several iterations [21]. By assigning colors to different heights, the system renders a topography that resembles a set of islands surrounded by an ocean. The peaks of the virtual landscape indicate abundant coverage on a particular topic, whereas valleys represent sparsely populated parts of the information space. For each peak, the most relevant keywords of documents in the vicinity are calculated and displayed as a caption. Users can select documents by clicking on an area of interest in the information landscape.

*Ontology View.* Ontologies retrieved from the extension process are serialized in OWL (Ontology Web Language) and then fed into the visualization framework, which uses *graphviz* ([www.graphviz.org](http://www.graphviz.org)) to determine the positions of individual nodes. The component includes hypernym and hyponym relations using arrows pointing to the more general concepts. It also represents all other relation types with a dashed line and a number indicating the relative strength of the association between the connected concepts based on



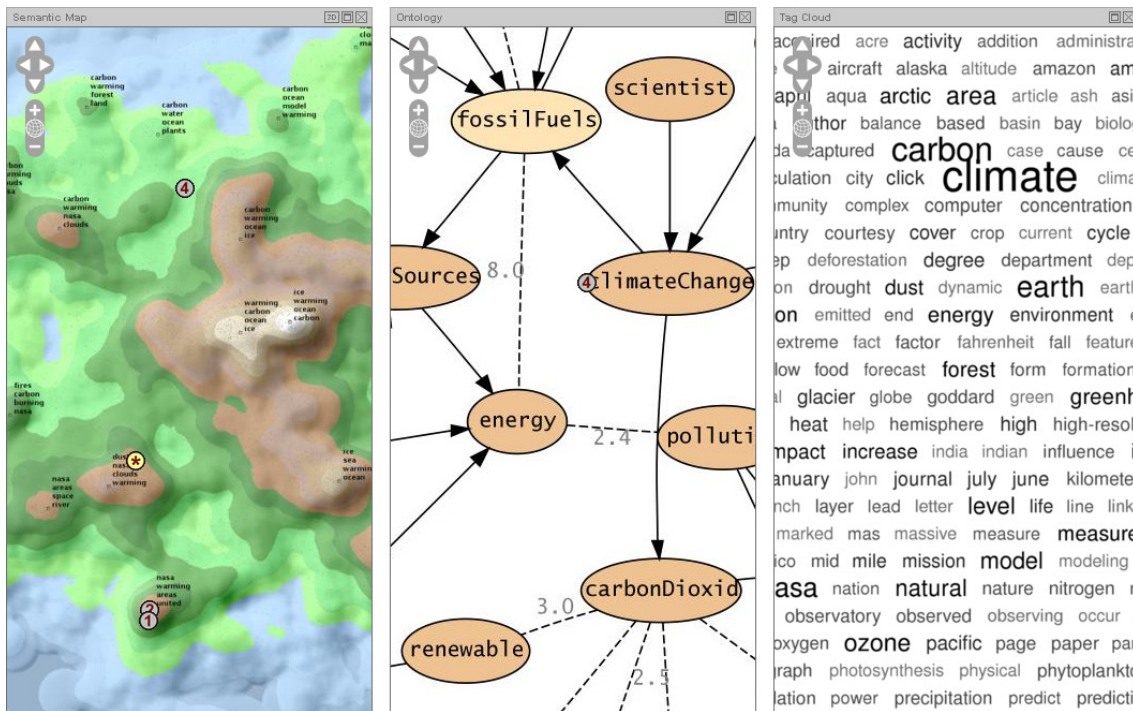


Figure 4: Visualizing semantic associations between concepts and documents via information landscapes, ontology graphs and tag clouds

the link assignment's spreading activation level. High values suggest a strong relation between the concepts, a value of 8 being the maximum due to the specific setup of the spreading activation network. The final visualization is then exported to the Web Map Server. Figure 4 shows a part of the domain ontology after applying the iterative extension process to approximately 80,000 documents mirrored from 50 news media sites in July 2007.

*Tag Cloud.* This view arranges significant terms alphabetically and positions them using a graph layout algorithm. The size and color of the terms reflect their specific importance within the corpus (i.e., their relative frequency). The component displays more frequent terms in black and less frequent terms in shades of gray using a smaller font. Users can select a topic of interest by clicking on a term, which triggers a full text search and displays a list of matching concordances.

#### 4.5 Temporal View

The temporal view summarizes major trends in current news media coverage. Each week, a set of keywords for the target corpus (= this week's documents) is calculated in relation to a static reference corpus. Attention (= relative frequency) and sentiment values for the five most popular keywords are calculated retroactively for the previous ten weeks.

Two buttons above the window allow the user to toggle between the relative frequency of the keywords over time and the news media sentiment towards these terms. Clicking on one of the keywords in the legend triggers a full text search. Figure 5, for example, reflects the economic and geopolitical events of August 2008 (rising oil prices and the call for new drilling sites; crisis in Georgia and news media coverage on the Georgian President Mikheil Saakashvili).

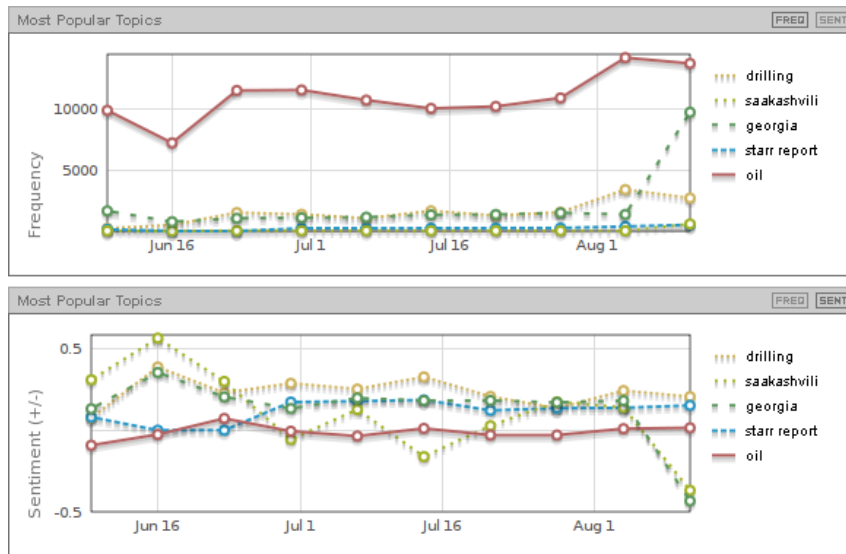


Figure 5: Relative frequency and sentiment of the five most popular keywords

## 5 Interoperability

Three types of machine-readable interfaces ensure the interoperability of the *Media Watch on Climate Change*: (i) *Web Services* offer access to important core components; (ii) *RSS feeds* deliver up-to-date information covering specific topics and locations; (iii) *microformats* enable a tighter integration with Web and desktop applications.

Web services that implement geotagging, significant phrase detection and keyword analysis as Tomcat Servlets (tomcat.apache.org) facilitate integration into third-party applications and allow combining multi-annotation service installations into service clusters for reasons of throughput and availability.

RSS feeds provide annotated document feeds of relevant topics and keywords. While the first feed delivers documents containing the *search tag(s)* in the specified sample, the other two feeds cover similar topics and nearby locations in relation to the *active document*. The feeds are available at the following addresses:

```
www.ecoresearch.net/climate/rss/tag/{sample}/{tag}
www.ecoresearch.net/climate/rss/topic/{sample}/{doc-id}
www.ecoresearch.net/climate/rss/location/{sample}/{doc-id}
```

Microformats such as hCard, hCalendar, and the XHTML Friends Network (XFN) provide straightforward techniques to embed semantically rich information into existing Web pages. Their markup is not as advanced and powerful as the Resource Description Framework (RDF), but aggregator services like *Technorati* (technorati.com) have demonstrated the usefulness of these tagging formats [24].

The *Media Watch on Climate Change* utilizes 'Geo' microformats to describe the target geography of the documents. 'Rel-tag' markup extends existing hyperlinks, marking the destination as a system-designated tag. Documents from the just-in-time information retrieval agents are annotated using the 'hReview' microformat, marking them as reviews including relevance ratings and review dates. Browser add-ons such as *Operator* (addons.mozilla.org/addon/4106) and *Tails Export* (addons.mozilla.org/addon/2240) provide a user-friendly interface for importing such entries into existing applications. Geo-metadata from the *Media Watch on Climate Change* can be presented as a link to mapping applications such as *Yahoo! Maps* (maps.yahoo.com) and *Google Maps* (maps.google.com), while 'rel-tag' markup refers users to community sites such as *Delicious* (www.delicious.com) and *flickr* (www.flickr.com).

## 6 Outlook and Conclusion

Automatically annotating content from heterogeneous sources creates knowledge repositories spanning multiple dimensions such as space, time and semantics. Advanced search methods should incorporate these dimensions and make them intuitively accessible. Multiple coordinated views are an effective means to address this challenge and complement relevance-ranked lists of documents, which still remain the dominant format to present the results of Web searches.

The *Media Watch on Climate Change* provides a prototype implementation of such an interface that visualizes contextualized information spaces through geographic maps, domain ontologies, information landscapes and tag clouds. Integrating these visualizations within a common framework supports information retrieval tasks. This paper describes the underlying methods and demonstrates that geospatial interface technology can serve as a generic image rendering engine to project various types of imagery. Future versions of the portal will provide 3D rendering of all supported visualizations through a common Java-based architecture.

Web-based user interfaces such as the *Media Watch on Climate Change* pave the way for a richer user experience and more active user participation, which can play an important role in improving semantic services [25]. Integrating social evidence sources will improve the quality of annotations and thereby lead to better search results. Follow-up research and user studies will investigate search performance by measuring average search times spent on pre-defined search tasks, investigate the portal's utility for searching and browsing as the two most common modes of interaction, and compare the usability of two-dimensional (e.g., information landscape) and three-dimensional (e.g., knowledge planet) visualizations.

### Acknowledgment

The *Media Watch on Climate Change* has been developed as part of IDIOM (Information Diffusion across Interactive Online Media; [www.idiom.at](http://www.idiom.at)) and RAVEN (Relation Analysis and Visualization for Evolving Networks; [www.modul.ac.at/nmt/raven](http://www.modul.ac.at/nmt/raven)), two research projects funded by the Austrian Ministry of Transport, Innovation & Technology (BMVIT) and the Austrian Research Promotion Agency (FFG) within the strategic objective FIT-IT ([www.fit-it.at](http://www.fit-it.at)). The current prototype has jointly been developed by Syed Kamran Ali Ahmad, Alexander Hubmann-Haidvogel, Walter Rafelsberger, Arno Scharl, Hermann Stern, Albert Weichselbraun, Gerhard Wohlgenannt and Dimitri Zibold. The authors wish to thank Arinya Eller for proof-reading the manuscript.

### References

- [1] A. I. Abdelmoty, P. D. Smart, C. B. Jones, G. Fu, D. Finch, A critical evaluation of ontology languages for geographic information retrieval on the internet, *Journal of Visual Languages and Computing* 16 (4) (2005) 331–358.
- [2] J. Allan, B. Carterette, J. Lewis, When will information retrieval be “good enough”?, in: *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*, ACM Press, New York, NY, USA, 2005, pp. 433–440.
- [3] E. Amitay, N. Har'El, R. Sivan, A. Soffer, Web-a-where: geotagging web content, in: *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '04)*, ACM Press, New York, NY, USA, 2004, pp. 273–280.
- [4] M. Q. W. Baldonado, A. Woodruff, A. Kuchinsky, Guidelines for using multiple views in information visualization, in: *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '00)*, ACM Press, New York, NY, USA, 2000, pp. 110–119.

- [5] D. Brodbeck, L. Girardin, Visualization of large-scale customer satisfaction surveys using a parallel coordinate tree, in: 9th IEEE Symposium on Information Visualization (InfoVis 2003), IEEE Computer Society, Seattle, USA, 2003, pp. 197–201.
- [6] S. Card, J. Mackinlay, B. Shneiderman, Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann, San Francisco, 1999.
- [7] R. Chimera, B. Shneiderman, An exploratory evaluation of three interfaces for browsing large hierarchical tables of contents, ACM Transactions on Information Systems 12 (4) (1994) 383–406.
- [8] T. Dunning, Accurate methods for the statistics of surprise and coincidence, Computational Linguistics 19 (1) (1993) 61–74.
- [9] C. Fellbaum, Wordnet an electronic lexical database, Computational Linguistics 25 (2) (1998) 292–296.
- [10] E. Gahleitner, W. Behrendt, J. Palkoska, E. Weippl, Knowledge sharing and reuse: On cooperatively creating dynamic ontologies, in: Proceedings of the 16th ACM Conference on Hypertext and Hypermedia (Hypertext-2005), ACM Press, Salzburg, Austria, 2005, pp. 208–210.
- [11] H. Joho, M. Sanderson, M. Beaulieu, A study of user interaction with a concept-based interactive query expansion support tool, in: Advances in Information Retrieval, 26th European Conference on Information Retrieval, 2004, pp. 42–56.
- [12] T. Kapler, W. Wright, Geotime information visualization, Information Visualization 4 (2) (2005) 136–146.
- [13] J. L. Leidner, An evaluation dataset for the toponym resolution task, Computers, Environment and Urban Systems 30 (2006) 400–417.
- [14] W. Liu, A. Weichselbraun, A. Scharl, E. Chang, Semi-automatic ontology extension using spreading activation, Journal of Universal Knowledge Management 0 (1) (2005) 50–58, [http://www.jukm.org/jukm\\_0\\_1/semi-automatic\\_ontology\\_extension](http://www.jukm.org/jukm_0_1/semi-automatic_ontology_extension).
- [15] G. Marchionini, Exploratory search: From finding to understanding, Communications of the ACM 49 (4) (2006) 41–46.
- [16] J. Nielsen, Multimedia and Hypertext: The Internet and Beyond, AP Professional, Boston, 1995.
- [17] K. Norman, L. Weldon, B. Shneiderman, Cognitive layouts of windows and multiple screens for user interface, International Journal of Man-Machine Studies 25 (1986) 229–248.
- [18] C. North, B. Shneiderman, Snap-together visualization: A user interface for coordinating visualizations via relational schemata, in: Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '00), ACM Press, Palermo, Italy, 2000, pp. 128–135.
- [19] B. Rhodes, P. Maes, Just-in-time information retrieval agents, IBM Systems Journal 39 (3/4) (2000) 685–702.
- [20] D. Roussinov, J. L. Zhao, Automatic discovery of similarity relationships through web mining, Decision Support Systems 35 (2003) 149–166.
- [21] V. Sabol, A. Scharl, Visualizing temporal-semantic relations in dynamic information landscapes, in: 11th International Conference on Geographic Information Science (AGILE-2008), Semantic Web Meets Geospatial Applications Workshop, AGILE Council, Girona, Spain.

- [22] A. Scharl, Tightly coupled geospatial interfaces for collaborative systems and just-in-time information retrieval agents, *Research in Computing Science (Special Issue: Semantics in Geoinformatics Applications)* 25 (2006) 3–18.
- [23] B. Shneiderman, P. Shafer, R. Simon, L. Weldon, Display strategies for program browsing: Concepts and an experiment, *IEEE Software* 3 (3) (1986) 7–15.
- [24] D. Sifry, The state of the live web, April 2007, Tech. rep., Technorati, Inc., <http://technorati.com/weblog/2007/04/328.html> (2007).
- [25] K. Siorpaes, M. Hepp, Games with a purpose for the semantic web, *IEEE Intelligent Systems & their Applications* 23 (2008) 50–60.
- [26] A. H. Turpin, F. Scholer, User performance versus precision measures for simple search tasks, in: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06)*, ACM Press, New York, NY, USA, 2006, pp. 11–18.
- [27] F. Yates, Contingency table involving small numbers and the  $\chi^2$  test, *Supplement to the Journal of the Royal Statistical Society* 1 (2) (1934) 217–235.